Replay 1(5) 2018
http://dx.doi.org/10.18778/2391-8551.05.01

Bartosz Maćkiewicz
Wojciech Mamak
Uniwersytet Warszawski

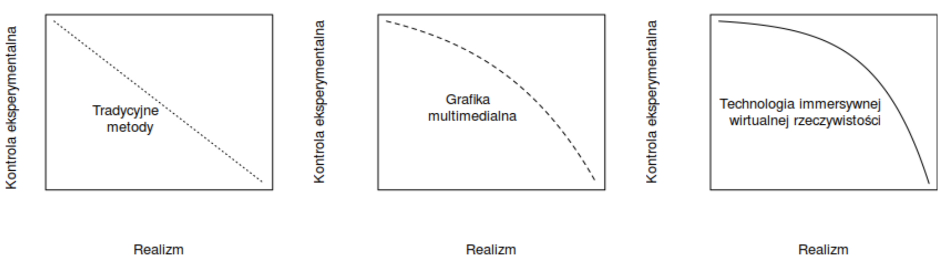
Gry jako moralne laboratorium. Gamifikacja dylematów moralnych bezzałogowych samochodów

W artykule pokażemy, jak gamifikacja schematów eksperymentalnych pozwala przedstawić w nowym świetle dylematy moralne związane z pojazdami bezzałogowymi. W pierwszej części dokonamy przeglądu kwestii związanych z wykorzystaniem gier komputerowych w badaniach eksperymentalnych. Rozważymy potencjalne pułapki czyhające na testujących intuicje etyczne z użyciem gier. Następnie omówimy pod względem obecności elementów gamifikacji zrealizowane do tej pory badania dotyczące społecznego postrzegania moralnych decyzji podejmowanych przez samochody bezzałogowe. Przedstawimy roboczą typologię gier ze względu na ich użyteczność jako narzędzi w badaniach nad różnymi typami problemów etycznych. Na tej podstawie proponujemy rozwiązania, jak unikać pułapek opisywanych w części pierwszej. Przedstawiamy trzy podstawowe podejścia do implementacji „modułu moralnego". Twierdzimy, że są one przekładalne na język klasycznych doktryn etycznych: utylitaryzmu, deontologii i etyki cnoty, przy zastosowaniu teorii etyki informacji. W ostatniej części tekstu stawiamy kilka otwartych problemów związanych z możliwością algorytmizacji i testowalności etyki. Tekst kończymy przeglądem możliwości, jakie dla tak rozumianej testowalności otwiera zastosowanie gier w schemacie eksperymentalnym.

Wykorzystanie gier w naukach eksperymentalnych

Gry komputerowe wykorzystywane były w naukach społecznych w zasadzie od pojawienia się gier automatowych w latach siedemdziesiątych. Szeroko stosuje się je np. w psychofizjologii (Järvelä et al., 2012), psychologii społecznej (Blascovitch et al., 2002) czy psychologii poznawczej (Washburn, 2003).

Zespół Blascovitcha (Blascovitch et al., 2002) zwrócił uwagę na konieczność kompromisu w kwestii schematu doświadczalnego w badaniach psychologii społecznej. Badacz posługujący się tradycyjnymi metodami, projektując badanie, musi wybrać między eksperymentalną kontrolą a realizmem. Jeżeli interesuje go maksymalna eksperymentalna kontrola, to wybierze metody jak najpełniej ją gwarantujące. W przypadku psychologii społecznej często będą to badania kwestionariuszowe, w których badani oceniają przedstawianą im sytuację. Metoda ta zapewnia dużą dozę kontroli, ponieważ badacz sam tworzy wszystkie szczegóły opisu danej sytuacji. Ocenianie opisanych w scenariuszach badawczych sytuacji niewiele ma jednak wspólnego z zachowaniem w rzeczywistych kontekstach społecznych. Jeżeli badacz chce zbliżyć do rzeczywistości schemat eksperymentalny, zamiast opisowych scenariuszy wybierze bardziej wyrafinowane metody (np. zaaranżowanie sytuacji z wykorzystaniem aktorów). To rozwiązanie gwarantuje dużą dozę realizmu, ze względu jednak na poziom skomplikowania i możliwość wpływu czynników zewnętrznych, poziom eksperymentalnej kontroli znacznie spada. Blascovitch i współpracownicy uważają, że chociaż zawsze mamy do czynienia z tego rodzaju kompromisem, to wykorzystanie gier komputerowych pozwala osiągnąć względnie duży poziom realizmu, nie oddając przy tym wiele z kontroli nad warunkami eksperymentalnymi.



W literaturze wskazuje się wiele zalet wykorzystywania gier w badaniach, dostrzegane są również ich wady. Katalog zalet rozpocząć należy od immersji. Zanurzenie w świat gry może sprawić, że badany zapomni o tym, że jest w sytuacji eksperymen-

talnej. Taka sytuacja zwiększa znacznie podatność na manipulacje eksperymentalne oraz poprawia jakość wykonywania zadań eksperymentalnych. Jeśli zadanie przedstawione jest w formie gry komputerowej, badani wykonują je szybciej, precyzyjniej i popełniają mniejszą liczbę błędów (Washburn, 2003). Wykorzystanie gier może mieć też pozytywne skutki dla trafności ekologicznej badania. Wysoki poziom realizmu oferowany przez gry komputerowe oraz zwiększone zaangażowanie badanego w eksperyment lepiej uzasadniają przełożenie otrzymanych wyników na sytuacje występujące w prawdziwym świecie.

Warto też zwrócić uwagę na obycie coraz większej liczby osób z grami komputerowymi: gra się już w 65% amerykańskich gospodarstw domowych (ESA, 2011), w Polsce gra 72% użytkowników internetu (Bobrowski, Rodzińska-Szary, Socha, 2015). Wbrew obiegowej opinii granie nie jest domeną jedynie najmłodszej grupy wiekowej. Badania przeprowadzone na polskich użytkownikach internetu, pokazują, że 25% użytkowników znajduje się w grupie wiekowej 35+. Gry mają bardzo wysoki współczynnik penetracji i dla badacza eksperymentalnego jest to duże ułatwienie. Złożone zadania eksperymentalne, mogą być nieprzystępne dla przeciętnego badanego lub wymagać skomplikowanego, czasochłonnego i kosztownego treningu. W przypadku gier komputerowych badani stykają się z medium dobrze sobie znanym i może to przyczynić się do lepszej jakości tak uzyskanych danych.

Wykorzystanie gier komputerowych nie jest jednak pozbawione wad. Po pierwsze, wiąże się z nimi jednak pewnego rodzaju brak kontroli. Gry są ze swojej istoty interaktywnym medium. Gracz może za pośrednictwem swojego awatara i mechanik gry wchodzić w interakcje z otoczeniem, a badacz może nie być w stanie przewidzieć wszystkich możliwości. W większości gier rozgrywka przynajmniej częściowo jest nieliniowa. Może to okazać się problemem w wypadku wykorzystania w badaniach eksperymentalnych gier komercyjnych (Järvelä et al., 2012). W badaniach eksperymentalnych bowiem bardzo często pełna kontrola nad bodźcami eksperymentalnymi jest warunkiem metodologicznej poprawności.

Drugą kwestią jest porównywalność wyników między uczestnikami badania. Z jednej strony, ze względu na wspomnianą wcześniej interaktywność gier, zagwarantowanie tego, że w przypadku każdego badanego bodziec eksperymentalny będzie taki sam bądź przynajmniej podobny. Z drugiej zaś strony wykorzystanie gier sprawia, że zadanie eksperymentalne, przed którym postawiony jest badany, może mieć inny charakter w zależności od jego wcześniejszej ekspozycji na ten gatunek rozrywki. W wypadku gier badani dysponują odmiennym zestawem umiejętności na różnym poziomie. Wytrawny gracz w *Counter-Strike'a* (Valve, 2000) jest w sta-

nie sprawniej poruszać się po przypominających labirynty planszach niż ktoś, kto nigdy wcześniej nie grał w produkcje z gatunku FPS. Szybkość i dokładność wykonywania zadania będą zależeć nie tylko od trudności ćwiczenia, indywidualnych własności danej osoby i wpływu manipulowanych przez eksperymentatora zmiennych, lecz również od jego czysto growej sprawności. W przypadku standardowych układów eksperymentalnych, w których nie wykorzystuje się gier, z reguły można założyć, że badani nie mieli wcześniej styczności z podobnymi zadaniami. Badania (Frey et al., 2007) wskazują, że różnica wynikająca z ekspozycji na gry może zostać zmniejszona, choć nie całkowicie zniwelowana, przez wprowadzenie do schematu eksperymentalnego części treningowej.

Najważniejszą z perspektywy naszego artykułu wadą wykorzystania gier komputerowych jest zagrożenie *powergamingiem*. Musimy liczyć się z tym, że subtelne wyzwania etyczne stawiane przez grę mogą zostać zignorowanie w imię maksymalizacji efektywności strategii. W znanym *Fable* (Lionhead Studios, 2004) możemy uzyskać najsilniejszy oręż dostępny w grze, jeżeli zamordujemy własną siostrę. To jednak, że ktoś dokonał tego wyboru, niekoniecznie musi oznaczać, że jest w rzeczywistości psychopatą. Osoba taka mogła zdystansować się od moralnego aspektu wyboru i dokonać morderstwa ze względu na wymierny w kategoriach mechaniki rozgrywki zysk.

Moral Machine

Wprowadzenie do powszechnego użycia bezzałogowych samochodów wiąże się z wieloma kontrowersjami. Pomimo że badania naukowe wskazują, że autonomiczne pojazdy pozwolą zmniejszyć liczbę ofiar ginących rokrocznie w wypadkach komunikacyjnych (Johansson & Nilsson, 2016), pewne kontrowersje dotyczące ich zachowania na drodze są przedmiotem publicznej debaty. Z naszej perspektywy szczególnie interesujący jest fakt, że problematykę społecznego postrzegania podejmowanych decyzji moralnych bada się za pomocą eksperymentów czerpiących pewne rozwiązania konstrukcyjne z gier. Przykładem takiego działania jest projekt *Moral Machine* prowadzony na MIT przez zespół Bonnefona, Shariffa i Rahwana. Projekt ten można uznać za kontynuację wcześniejszych badań tego zespołu (Bonnefon, Shariff, Rahwan, 2016). Autorów szczególnie interesowała kwestia w jaki sposób powinien zachowywać się bezzałogowy samochód w sytuacjach, w których zagrożone jest życie użytkownika pojazdu lub innego uczestnika ruchu. Czy ludzie uważają, że bezzałogowe auto powinno poświęcić życie pasażerów, aby uratować przechodnia lub pieszego? Okazuje się, że 76% respondentów sądzi, że pożądane

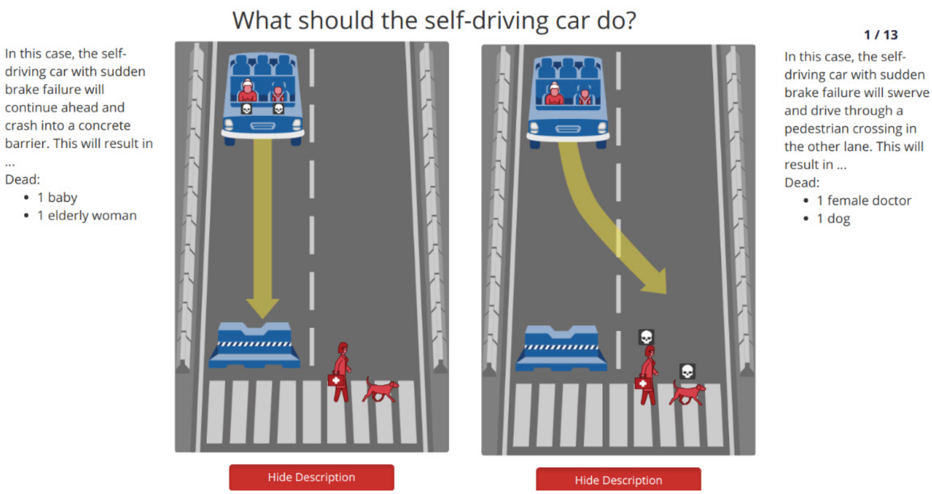
jest poświęcenie jednego pasażera pojazdu po to, aby uratować dziesięciu przechodniów. U badanych widać preferencje utylitarystyczne. Liczba ocalonych, dzięki poświęceniu pasażera, żyć jest silnie skorelowana z moralną aprobatą takiej ofiary. Badacze manipulowali również statusem pasażera. Wyniki były bardzo podobne, nawet gdy w scenariuszu pasażerem, którego autonomiczny pojazd miał poświęcić, był członek rodziny badanego. Aprobata takiego poświęcenia była słabsza niż w wypadku, gdy status osoby siedzącej w samochodzie był nieokreślony, ale dalej obecny był utylitarystyczny sposób myślenia.

Choć ludzie uważają, że pojazdy powinny postępować utylitarnie, to jak wynika z omawianego eksperymentu, mając do wyboru kupno samochodu lub pojazdu zaprogramowanego tak, by za wszelką cenę chronił pasażerów, wybraliby ten drugi. Co więcej, badani w większości nie zgadzają się na wprowadzanie jakichkolwiek regulacji wymuszających implementację w pojazdach utylitarystycznych algorytmów. Ludzie są więc niemal jednomyślni co do tego jak samochody powinny reagować w kryzysowych sytuacjach, nie chcą jednak, by ich własne auta były tak zaprogramowane.

Jak wspomnieliśmy, zespół Bonnefona kontynuuje badania nad moralnością[1] autonomicznych pojazdów za pomocą portalu *Moral Machine* (http://moralmachine.mit.edu). Ich celem jest zrozumienie tego, jak ludzie podejmują decyzje moralne oraz jak postrzegają dokonujące je maszyny. Dzięki temu chcą stworzyć *crowdsource'owy* obraz tego, jak powinny postępować maszyny stojące w obliczu dylematu moralnego. Na portalu można wcielić się w rolę sędziego oceniającego przypadki, w których pojazd bezzałogowy musi dokonać wyboru. Za każdym razem można określić jedno z dwóch zachowań – ratowanie pasażerów lub ratowanie przechodniów i pieszych. Sytuacje, które ocenia użytkownik, różnią się pod wieloma względami: liczbą zagrożonych osób, ich płcią, wiekiem i statusem społecznym, a nawet rodzajem sylwetki (por. rys. 1). W niektórych scenariuszach przechodnie nie stosują się do przepisów ruchu drogowego.

---

1 Zdefiniowanie moralności wykracza poza zakres niniejszej pracy, jest to bowiem problem trapiący filozofów od tysiącleci. Niewątpliwie jednak nasz aparat poznawczy pozwala na intuicyjne rozróżnianie norm moralnych od norm pozamoralnych. Psychologowie rozwojowi wskazują, że umiejętność ta występuje we wczesnym wieku wśród przedstawicieli wszystkich kultur. Rozległe omówienie tej problematyki można znaleźć w (Huebner et al., 2010). Argumentacja przedstawiona w pracy nie opiera się na możliwości precyzyjnej klasyfikacji przypadków granicznych, dlatego intuicyjne kryterium jest zupełnie wystarczające.



Rys. 1. Przykładowa sytuacja oraz interfejs aplikacji *Moral Machine*

Po ocenieniu odpowiedniej liczby wypadków, każdemu z użytkowników wyświetla się podsumowanie jego wyborów. Swoją formą przypomina „tablice wyników" przedstawiane w grach *online* po wygranym meczu. Użytkownik widzi jak jego preferencje moralne wypadają na tle innych odbiorców oraz czy liczba uratowanych osób lub ich płeć są czynnikami silniej wpływającymi na jego decyzje niż ma to miejsce w populacji. Dowiaduje się, którą z osób występującą w scenariuszach zabijał najczęściej (*Most Killed Character*), a którą najczęściej ratował (*Most Saved Character*). Otrzymanymi informacjami może podzielić się na portalach społecznościowych, np. na *Facebooku*. Dodatkowo ma szansę tworzyć własne scenariusze i dzielić się nimi jak w *LittleBigPlanet* (Media Molecule, 2008). Dzięki tym „growym" elementom badanie zyskało ogromną popularność. Jak do tej pory nie ukazała się jednak żadna publikacja podsumowująca wyniki tego eksperymentu.

Nie jest jasne, czy tak pozyskane dane są bardziej wiarygodne niż te, które zebrać można za pomocą tradycyjnego eksperymentu. Część użytkowników mogła „przeklikać" kwestionariusz, nie zastanawiając się nad ocenianymi sytuacjami. Atrakcyjna forma badania może dawać nadzieję, że znaczna część respondentów poważnie podeszła do zadania. Istnieją też statystyczne techniki odfiltrowania tych odpowiedzi, które nie wykazują odpowiedniego poziomu wewnętrznej spójności (DeSimone et al., 2015).

Pod pewnymi względami podobne tematycznie analizy przeprowadził zespół Sütfelda (2017), wykorzystując zestaw do wirtualnej rzeczywistości, aby „posadzić" badanych na miejscu kierującego pojazdem w momencie etycznego dylematu. Osoby biorące udział w eksperymencie musiały wybierać między różnego rodzaju przeszkodami (przedmiotami nieożywionymi, zwierzętami, ludźmi występującymi pojedynczo lub w grupach). Okazało się, że decydujący jest bilans „wartości życia" (*life-value*) porównywanych ofiar. W przypadku ludzi koreluje ona silnie z wiekiem zagrożonej osoby[2]. Wydaje się przy tym, że inne czynniki nie były brane pod uwagę, a badani w wyborach konsekwentnie kierują się wiekiem. W przypadku skróconego czasu reakcji osłabia się tendencja do ważenia wyborów według „wartości życia". Zdaniem autorów jest to argument za tym, że ludzie mają zdefiniowane preferencje dotyczące wyborów, jakie skłonni byliby podjąć. Zaobserwowana różnica w wynikach jest ich zdaniem konsewkencją większej roli błędów. Jest to mocny argument za implementacją automatycznych algorytmów decyzyjnych. Automatyzacja taka mogłaby wyeliminować możliwość podjęcia przez kierowcę błędnej decyzji.

Wyzwania badań moralności za pomocą gier

W kontekście badań nad moralnością można wyróżnić trzy rodzaje gier. Po pierwsze, mamy gamifikowane układy eksperymentalne specjalnie stworzone dla potrzeb badawczych. Charakteryzują się małą interaktywnością i relatywną prostotą. Przykładem jest serwis *Moral Machine* lub eksperymenty przeprowadzane w środowisku wirtualnym (badanie zespołu Sütfelda). Drugi typ to gry z mocnym nastawieniem na fabułę, liniowe i oskryptowane z kilkoma wyborami moralnymi osadzonymi w ich narracji. Paradygmatycznym przykładem jest seria *The Walking Dead* (Telltale, 2012). Trzecim typem są gry z otwartym światem i swobodą interakcji z elementami środowiska: w wypadku rozgrywki dla jednego gracza to np. *Fallout 3* (Bethesda, 2008), w wypadku rozgrywki wieloosobowej – *Minecraft* (Mojang, 2011) i *Eve Online* (CPP Games, 2003). W tym rodzaju gier użytkownicy mogą do pewnego stopnia swobodnie kształtować swoje otoczenie.

Przed osobami chcącymi wykorzystać gry komputerowe lub ich elementy do badania moralności stoją pewne wyzwania specyficzne dla domeny moralnej. Po pierwsze, muszą zdecydować jak zaprezentować dylematy etyczne w taki sposób, by miały dla gracza znaczenie. Oznacza to, że wybory muszą mieć odpowiednią wagę; powinny być dla odbiorcy angażujące i rozpoznawane jako decyzje moralne, nie zaś jako środki

---

2 Potwierdza to wcześniejsze wyniki uzyskane przez Johansson-Stenman i Martinssona (2008).

do osiągnięcia sukcesu w grze. Podczas wyboru musi być również jasne, że decyzja ma istotne etycznie skutki – fakt ten nie może być ukrywany przed graczem, który powinien mieć przynajmniej podstawowy zasób informacji pozwalający mu na rozważenie danej sytuacji (Sicart, 2013).

Po drugie, należy zachęcić graczy do zaangażowania się w dylematy moralne. Ważne jest tu zaimplementowanie moralności jako elementu mechaniki gry. Z naszej perspektywy złym przykładem takiego rozwiązania jest przywoływana już seria *Fable*, w której wybory moralne są centralnym elementem mechaniki gry. Decyzje te wpływają na wygląd bohatera – jeżeli będziemy postępować niemoralnie, protagoniście wyrastają rogi. W rzeczywistości gracz nie dokonuje żadnego istotnego z punktu widzenia moralnego wyboru, skoro określa jedynie, czy chce być czarnym charakterem, czy też woli stanąć po stronie dobra (Sicart, 2009). Trywializuje to problem podejmowania etycznych wyborów. Wyzwania, przed którymi stoimy, mają znacznie subtelniejszy charakter i często polegają na ważeniu różnych dóbr, potencjalnych skutków działań oraz ryzyka z nimi związanego. Czarno-białe systemy decyzji moralnych o precyzyjnie określonych i z góry znanych konsekwencjach, nie odwzorowują ludzkiej kondycji moralnej.

Po trzecie, gracze nie powinni optymalizować swoich strategii pod kątem rozgrywki. *Powergaming* jest zagrożeniem dla każdego projektu gry, która ma nas angażować w rozważania etyczne. W wielu grach pewien rodzaj decyzji moralnych ma lepsze skutki niż inny. Seria *Baldur's Gate* (Bioware, 1998; 2000) znana jest tego, że za pomocą wartościowych przedmiotów otrzymywanych za wykonywanie zadań, nagradza graczy stojących po stronie dobra. Podobne obawy budzą gry wieloosobowe, w których istotnym elementem jest rywalizacja. Jeżeli pewien rodzaj decyzji moralnych będzie dawał wymierną przewagę nad innymi użytkownikami, to odbiorcy będą dokonywać wyborów tylko ze względu na tę korzyść.

Konkurencyjne podejścia do konstrukcji samochodów bezzałogowych

Pojazdy bezzałogowe wyposażone są w moduł samosterujący obejmujący mechanizm podejmowania decyzji. Samochód może przyjmować na siebie albo wybraną część obowiązków związanych z jazdą i manewrami (*automated* – automatyzowane), bądź dążyć do całkowitej autonomiczności (*autonomous* – samodzielne). Amerykańska agencja do spraw bezpieczeństwa ruchu drogowego opisuje to spektrum pięciostopniową skalą (NHTSA, 2013), gdzie 0 oznacza ograniczenie podmiotowości pojazdu do wydawania ostrzeżeń. Powszechnie dostępne samochody są już wyposażone w systemy automatycznego zwalniania przed pojazdem na wprost siebie, oferują utrzymy-

wanie się w pasie jazdy oraz samodzielnie parkują (poziom 1). Kolejnym etapem będącym już w stałej ofercie, są systemy autopilotów[3]. Samochody wyposażone w takie rozwiązanie wykonują w całości zadania kierowcy, działają jednak tylko pod nadzorem człowieka, który w razie kryzysowej sytuacji może przejąć stery. Poziomy 2, 3 oraz 4 opisują wzrastającą, lecz wciąż ograniczoną „odpowiedzialność" samochodu. Poziom piąty to pełna autonomia, gdzie człowiek wyznacza jedynie punkt docelowy.

Pojazdy bezzałogowe mogą się różnić modułem moralnym, tzn. częścią oprogramowania, która odpowiada za podejmowanie decyzji moralnych. Istnieje tu całe spektrum podejść. „Kod moralny" można zaprojektować w sposób właściwy tradycyjnej sztucznej inteligencji, tj. całkowicie i *explicite* algorytmicznie. Moduł taki byłby wyposażony w zestaw przesłanek reprezentujących dyrektywy zachowania i wartości oraz w prerogatywy moralne i reguły inferencje pozwalające na podejmowanie decyzji w konkretnych sytuacjach.

Na drugim krańcu spektrum znajduje się wykorzystanie sieci neuronowych lub innych technik uczenia maszynowego. Przy takim podejściu godzimy się na element konstrukcyjnej „czarnej skrzynki". Pojęcie to nawiązuje do postulowanej w psychologii behawioralnej idei nieprzejrzystości pewnych procesów w systemie poznawczym i ograniczenia się w jego opisie tylko do materiału wejściowego (tu: sytuacji, które napotka samochód) i materiału wyjściowego (podjętych decyzji). Gdy architektura poznawcza systemu nie ma wbudowanych reguł, tak jak rozwiązania oparte na sieciach neuronowych, sieci te muszą zostać wytrenowane. Uczenie to może odbywać się pod nadzorem (*supervised learning*) lub też bez nadzoru (*unsupervised learning*). W pierwszym wypadku trening polega na „karmieniu" sieci możliwie dużym zestawem przykładowego materiału wejściowego adekwatnego dla pożądanego zadania i ocenie operacji wykonanych przez maszynę. Następnie dostosowuje się parametry neuronów tak, by zmaksymalizować skuteczność sieci w wykonywaniu określonego zadania. W przypadku modułu moralnego bezzałogowych samochodów danymi treningowymi, z których uczyłaby się sieć, musiałyby być konkretne problematyczne sytuacje opisane za pomocą pewnego zestawu własności oraz ich rozwiązania dokonane przez człowieka. Na podstawie takich danych sieć uczyłaby się naśladować ludzi w podejmowaniu moralnych decyzji.

W wypadku ćwiczeń bez nadzoru zestaw treningowy nie jest właściwie oznaczony, pojedynczym przypadkom (zadaniom, scenariuszom) nie towarzyszy odpowiedź,

---

3 Przykładem takich rozwiązań są oferowany przez Toyotę system „Anioła stróża" (*Guardian Angel*) oraz montowany w Teslach „Autopilot".

z którą system ma porównać swoją decyzję. Rozwiązanie takie sprawdza się, gdy w danym zestawie chcemy rozpoznać pewne prawidłowości lub sprawić, by system klasyfikował zadane sytuacje według wybranych przez siebie cech, a nie z pomocą wytycznych reprezentowanych w uczeniu nadzorowanym przez badacza. W tym sensie nauka z nadzorem jest rodzajem drogi pośredniej między tradycyjnym algorytmicznym kodowaniem opartym na regułach SI a uczeniem maszynowym bez nadzoru.

Rozróżnienia te dotykają istoty implementowania oceny i wykonywania działań etycznych u maszyn. Różnice te można wyłożyć z użyciem klasycznych paradygmatów etycznych: deontologicznego, utylitarystycznego i etyki cnoty[4].

Uważamy, że można wskazać odpowiedniość między inżynieryjnymi podejściami do problemu moralnych maszyn a koncepcjami etycznymi. Można wskazać elementy wspólne deontologicznej tradycji etycznej i klasycznego podejścia do sztucznej inteligencji oraz podejścia opartego na uczeniu maszynowym i etyki cnoty. Koncepcje utylitarystyczne plasują się między tymi dwoma podejściami. Odpowiedniość ta jest istotna zarówno z punktu widzenia teoretycznego, jak i praktycznego dotyczącego budowy rzeczywistych bezzałogowych pojazdów.

Podejście klasyczne zakłada, że decyzja podejmowana jest na podstawie *explicite* formułowanych reguł, których system trzyma się ściśle. Zasadami tego typu mogą być dyrektywy zachowania i rozwiązywania dylematów, w których jasno określono właściwy schemat działania. Można to traktować jako ucieleśnienie deontologicznej wizji etyki jako powinności podległej pewnym nieusuwalnym i niepodważalnym zasadom[5]. W wypadku samochodów autonomicznych, zestaw reguł odpowiedzialnych za rozwiązywanie dylematów bierze się ze świadomego wprowadzenia go przez programistę. Tak zaprojektowana maszyna byłaby praktyczną implementacją etyki deontologicznej.

Podobnie sprawa wygląda z pewnymi wersjami podejścia opierającego się na uczeniu maszynowym i utylitaryzmem zakładającym w najprostszym wariancie, że o moralnej wartości zachowania decyduje to, czy maksymalizuje ono sumę szczęścia[6]. Praktycznym problemem zastosowania tej koncepcji jest trudność oceny, jak bardzo

---

4 Nawiązujemy tutaj do koncepcji Wendella Wallacha i Colina Allena (2008), którzy scharakteryzowali dwa podejścia do moralności maszyn: wstępujące (*bottom-up*) oraz zstępujące (*top-down*) i powiązali je z istniejącymi koncepcjami etycznymi.

5 Obszerne omówienie klasycznych i współczesnych koncepcji deontologicznych znaleźć można w (Alexander, Moore, 2016)

6 Utylitaryzm jest jednym rodzajem koncepcji konsekwencjalistycznej, to znaczy zakładającej, że głównie (bądź jedynie) skutki czynów mają znaczenie moralne. Omówienie ogólnych założeń konsekwencjalizmu znaleźć można w (Sinnott-Armstrong, 2015).

zmniejszy lub zwiększy się suma szczęścia na świecie w wyniku określonego działania. Eksperyment, taki jak opisany wcześniej *Moral Machine*, może nam dostarczyć informacji, jaką wagę w sytuacjach problemowych ludzie przypisują różnym czynnikom przy określaniu bilansu strat i zysków. Takie dane pojazd bezzałogowy mógłby potraktować jako podstawę do przeprowadzenia utylitarystycznego rachunku.

Trzecia z rozważanych alternatyw konstrukcyjnych – uczenie głębokie (*deep learning*) – znajduje podbudowę w etyce cnoty. Ta wywodząca się od Arystotelesa koncepcja głosi, że dobre moralnie są te czyny, które wypływają z odpowiednio ukształtowanego charakteru[7]. Jedną z cech odpowiednio ukształtowanego charakteru jest mądrość praktyczna (*phronesis*), będącą cnotą, z której wypływają dobre wybory i której jedynym sposobem wykształcenia jest nauka i naśladowanie moralnych wzorów. Umiejętności tej nie da się sprowadzić do podążania za ustalonymi regułami. Podobnie możemy myśleć o sieci neuronowej uczącej się podejmować decyzje – wykształca cnotę *phronesis*.

Te trzy alternatywy można także rozpatrywać z punktu widzenia ich oddolności (*bottom-up*) lub odgórności (*top-down*) w kontekście etyki informacji (Floridi, 2013). Teoria ta zakłada, że wyzwania, przed jakimi stawia nas rozwój technologii, sprawiają, że tradycyjne doktryny przestają być samodzielnie wydolne i wymagają uzupełnienie bądź zastąpienia. Zwolennicy etyki informacji uważają, że niewłaściwe spojrzenie na niektóre problemy etyczne bierze się z braku akceptacji tego, że niektóre wybory moralne są u swej podstawy problemami związanymi z przetwarzaniem informacji. Tezę tę można głosić w dwóch wersjach. W jednej z nich można uważać, że etyka informacji ma odgrywać rolę alternatywy dla klasycznych doktryn, ze względu na ich niedostosowanie do współczesnych wyzwań. Twierdzić się będzie wtedy, że klasyczne systemy etyczne nie oferują właściwych wskazań ze względu na wkroczenie na arenę rozważań podmiotów moralnych takich jak komputery, roboty, czy właśnie AV oraz pojawianie się zjawiska rozpraszania odpowiedzialności (jak w przypadku relacji między programistą, kontrolerem i prawodawcą a urządzeniem o jakimś stopniu autonomiczności). Traktowanie niektórych decyzji etycznych jako zadań wykonywanych przez dowolne procesory informacji lub całe ich sieci bez przesądzania o sposobie ich fizycznej realizacji pozwala ominąć te przeszkody.

---

7 Współczesna etyka cnoty nie jest jednak wyłącznie rozwijaniem koncepcji Arystotelesa. W swoich dociekaniach posuwa się znacznie dalej, starając ustosunkować się do współczesnych dokonań psychologii moralnej i społecznej. Omówienie historycznych koncepcji etyki cnoty znaleźć można w (Hursthouse, Pettigrove, 2016).

Można jednak uważać, że etyka informacji jest rozszerzeniem klasycznych rozważań etycznych — rodzajem „teorii parasolowej" (*overarching scheme*). Nasza interpretacja różnych sposobów konstruowania modułów moralnych, które są mechanizmami przetwarzania informacji, jest przykładem pójścia właśnie tą drugą drogą. Przy takim rozwiązaniu możemy adekwatnie uchwycić różnice pomiędzy wciąż stosowalnymi doktrynami klasycznymi, jednocześnie nadając im uwspółcześnione odczytanie. Dzięki temu możemy także porównywać te rozwiązania z punktu widzenia zasygnalizowanych już kategorii oddolności i odgórności. Przetwarzanie odgórne charakteryzuje się uwzględnieniem, przynajmniej jako część informacji wejściowej, informacji pochodzących z wyższych pięter systemu. Wsad ten ma często formę pre-programowanych reguł działania zakodowanych w systemie od początku i wykorzystywanych do przetwarzania otrzymywanych danych na niższych poziomach. Przetwarzanie oddolne natomiast zakłada, że wyższe piętra są efektem przetwarzania informacji z niższych pięter, jednak nie odwrotnie. Informacje otrzymywane i przetwarzane na niskich poziomach (np. detekcyjnych/percepcyjnych) są przekazywane wyżej i tam interpretowane (lub też wykorzystywane do aktualizacji wyższych poziomów). Sama praca niższych poziomów obywa się jednak bez ingerencji wyższych pięter. Reguły nie są systemowi z góry dane, mogą co najwyżej pojawić się pod wpływem dostosowywania się do sygnału wejściowego (np. poprzez ustalanie się wag w sieciach koneksjonistycznych).

Łatwo zauważyć, że propozycje klasycznie algorytmiczne, odpowiadające etyce Kantowskiej lub podobnym oparte są przeważnie na przetwarzaniu odgórnym, a uczenie maszynowe przeciwnie – na oddolnym. Utylitarystyczny stan pośredni stawiający na naśladowanie panujących społecznych preferencji etycznych z pomocą ograniczonego zestawu zasad zwierzchnich, kierujących nauką tejże mimikry, staje się przy takim postawieniu sprawy połączeniem obu tych rodzajów przetwarzania. Postulowany przez nas element gamifikacji w zbieraniu danych odpowiadałby za wstępujący komponent tego schematu.

Algorytmizacja etyki

Koronnymi argumentami za zastosowaniem algorytmów do podejmowania decyzji są przede wszystkim efektywność i bezstronność. Wedle algorytmizacyjnej obietnicy program wykonuje pożądane zadanie, a wszelkie podejmowane po drodze decyzje są efektem poprawnego wnioskowania, obiektywnej procedury niezależnej od niepożądanych uprzedzeń, preferencji, emocji czy celów konkretnych jednostek lub grup interesów. Algorytmizacja, bazując na ścisłym definiowaniu obiektów swojego

działania i reguł postępowania, zdaje się też obiecywać zwiększenie transparentności w podejmowaniu decyzji. Należy jednak wystrzegać się bezkrytycznej wiary w pełną realizację tych obietnic. Jak wskazują teoretycy algorytmów (Mittelstadt et al., 2016), całkowicie wolny od uprzedzeń (*bias-free*), obiektywny, transparentny algorytm oferujący neutralne i niepodważalne odpowiedzi, jest mrzonką. Widać to szczególnie wyraźnie w przypadku algorytmów moralnych. Dzieje się tak, ponieważ w algorytmy nieuchronnie wbudowana jest niepewność (częściowa), nieprzejrzystość oraz normatywna niekonkluzywność. Jak wskazują Illari i Russo (2014), każdy algorytm jest tak dobry, jak dane, na których operuje, a każdy zestaw danych jest obdarzony immanentną skazą (np. niepełnej reprezentatywności lub częściowego braku wiarygodności). Parafrazując hasło Gitelman (2013): „surowe dane są oksymoronem". Innym problemem związanym z wbudowaną w system niepewnością jest brak nadziei na ustanowienie twardych związków przyczynowych na podstawie zestawów danych. Klasyczne argumenty Humeowskie pozostają w mocy, a nawet zyskują na sile, gdy podstawą wyprowadzanych wniosków jest obfitujący w nieregularności (anomalie) duży pakiet danych. Aby ominąć te problemy, przyjmuje się pojęcie „rozpoznania wystarczającego do działania" (*actionable insight*), które można określić jako próbę odniesienia się do rodzaju „roboczej kauzalności" czy też „efektywnej kauzalności" bez stosowania pojęcia „kauzalność". Trudno jednak postrzegać to jako cokolwiek innego niż unik.

Nieprzejrzystość jest kolejnym problemem związanym z algorytmizacją decyzji moralnych (Miller i Record, 2013). Badacze zwracają uwagę na fakt, że wiele algorytmów cechuje się brakiem transparentności w stosunku do swoich użytkowników. Twierdzą, że sytuacja taka – za której przykład stawiają działanie internetowych wyszukiwarek – może obniżać wartość epistemiczną wiedzy uzyskiwanej za ich pośrednictwem, a co za tym idzie, osłabiać stopień racjonalnego uzasadnienia decyzji podejmowanych na ich podstawie. Wiele algorytmów ma charakter komercyjny i są chronione przed wglądem podmiotów zewnętrznych, w tym użytkowników.

Problem nieprzejrzystości osiąga głębszy poziom w przypadku algorytmów opartych częściowo lub całkowicie na uczeniu maszynowym. Jak wspominaliśmy, w przypadku działań takich jak spontaniczne kategoryzacje, klastrowanie czy podejmowanie decyzji – od programów uczących się grać w gry, do programów uczących się zachowania na drodze, trudno jest orzec, jak w zasadzie została podjęta dana decyzja. Pytania o podstawę dla danego wyboru, czy możliwość zrozumienia konkretnej decyzji przez człowieka, nie mają tutaj charakteru trudności potencjalnie usuwalnej. Sytuacja taka ma miejsce chociażby w przypadku braku chęci czy możliwości ze stro-

ny użytkowników, aby dowiedzieć się, na jakiej zasadzie działają programy filtrujące informacje, z których korzystają (np. problem baniek informacyjnych w mediach społecznościowych) lub niemożliwości mającej charakter blokady prawnej (przypadek ochrony „arkanów sztuki" przez producentów wyszukiwarek). W przypadku uczenia maszynowego mamy jednak do czynienia z trudnością nieusuwalną, natury epistemologicznej *sensu stricto*.

Wreszcie potencjalnym problemem, szczególnie łatwo zauważalnym w przypadku pojazdów bezzałogowych, jest kwestia normatywnej niekonkluzywności. Wyobraźmy sobie, że skonstruowano mechanizm, który w sposób przejrzysty i powszechnie akceptowalny podejmuje decyzje zgodne z oczekiwaniami społeczności, w ramach której ma operować. Założenia wbudowane w system uznane są za rozsądne, reguły wnioskowania za dobrze zdefiniowane, a dyrektywy rozwiązywania dylematów za tożsame z intuicjami moralnymi ludzi. System ten został doprowadzony do wysokiej sprawności dzięki konsekwentnej nauce reagowania w wielkiej liczbie sytuacji, co do których ustalono pożądany sposób zachowania, na przykład na podstawie eksperymentów pokrewnych projektowi *Moral Machine* lub – lepiej – jakiejś jej zgamifikowanej formy. Nieuchronnie jednak maszyna taka prędzej czy później znajdzie się w sytuacji dotychczas nieznanej. Załóżmy, że maszyna podejmuje decyzję, która w powszechnej opinii uważana jest za nieintuicyjną lub wręcz błędną. Kto w takiej sytuacji ma rację? Czy powiemy, że algorytm popełnił błąd? Można sobie wyobrazić, że taki nieoczekiwany i niepożądany wynik osiągnięty został w ramach zaprogramowanego lub wyuczonego funkcjonowania. Czy jest to podstawa do wprowadzania zmian w algorytmie? Czy nie można twierdzić, że skoro maszyna działała w zgodzie z zaakceptowanymi pierwotnie przesłankami i regułami inferencji, to ona „ma rację" i wskazuje nam właściwą ocenę sytuacji? Czy można powiedzieć, że w tym sensie maszyna dokonała rodzaju „etycznego odkrycia"? Sądzimy, że postawienie tych pytań jest nieodzowne dla podjęcia rozsądnej dyskusji o prawdziwych konsekwencjach testowalności etyki i możliwości jej programowalności.

Co można testować używając gier? Podsumowanie

Wykazaliśmy, że testowanie intuicji etycznych z użyciem gier jest z perspektywy etyki informacji podejściem oddolnym. Za pomocą gier możemy testować zarówno sposoby podejmowania decyzji w określonym rodzaju ściśle zdefiniowanych sytuacji (studia przypadków, gry tworzone specjalnie na ten użytek), jak i złożone interakcje w obrębie rozbudowanego świata, potencjalnie zasiedlonego przez wielu ludzkich graczy. Balansowanie między tymi dwiema tendencjami pozwala zawrzeć pożądany

kompromis między kontrolą eksperymentalną a ekologiczną trafnością uzyskiwanych danych. Zaletą testowania z użyciem gier jest duża ilość różnorodnych danych możliwych do pozyskania w łatwy sposób.

Kwestia różnorodności danych jest zasadnicza dla ich reprezentatywności i skuteczności. Dla efektywnej nauki sieci neuronowej rozpoznawania obrazków kotów, pożądane jest przedstawienie zestawu treningowego przedstawiającego możliwie różnorodne przypadki tego, co chcemy zaklasyfikować jako „kota". Podobnie jest dla decyzji moralnych – im bardziej różnorodne są przypadki, z którymi zetknie się sieć, tym bardziej wiarygodne będzie „pojęcie" moralności, jakie wypracuje w procesie treningu. Dla niektórych efektów posłużyć się także można danymi uzyskiwanymi nie od badanych specjalnie w tym celu, ale przez analizę zachowania graczy grających dla przyjemności bądź rywalizacji. Duża ilość danych wiąże się z ich potencjalną różnorodnością. Aspekt ten nie jest jednak rzecz jasna specyficzny dla gier, ale informacje takie jak szybkość podejmowania decyzji czy dane psychofizjologiczne (tętno, aktywność mózgowa) mogą być przedmiotem analizy i są w ten sposób wykorzystywane (przykładem jest badanie nad wpływem umiejętności gry w Quake'a III Arena na sprawność psychomotoryczną – Frey, et al., 2007). Dane te mogą również służyć do oceny tego, jakie systemy poznawcze były zaangażowane w analizę sytuacji i podejmowaną decyzję lub jak stresującym doświadczeniem była konfliktowa sytuacja moralna. Różne parametry psychofizjologiczne stanowią wskaźnik obciążenia poznawczego, co pozwala modelować mechanizmy podejmowania decyzji moralnych. Kilka badań wskazało na różnice w rozkładzie podejmowanych decyzji w zależności od tego, czy badani mieli do dyspozycji mało czasu na decyzję lub wystarczająco dużo do namysłu (Sutter et al., 2011; Paxton et al., 2012; Sutfeld et al., 2017). Zdaniem niektórych, świadczy to o prawdziwości teorii „podwójnego przetwarzania" (*dual process theory*) (Greene, 2004; Cushman, 2013), zakładającej, że w podejmowaniu decyzji moralnej mogą uczestniczyć dwa odmienne systemy poznawcze. Jeden odpowiedzialny jest za szybkie, emocjonalne rozpoznawanie sytuacji, drugi zaś nastawiony jest na rozumowanie i analizę sytuacji, faworyzując rozwiązania zgodne z rachunkiem utylitarystycznym.

Sądzimy, że stosowanie schematów gamifikacyjnych jako symulacji realnych wyborów o dużej immersyjności doświadczenia dla badanego przy właściwym projektowaniu eksperymentów, z pominięciem potencjalnych trudności, omówionych w pierwszej części tekstu pozwala maksymalizować rzetelność zgromadzonych danych, upodabniając możliwie decyzję do tej podejmowanej w świecie rzeczywistym, zachowując przy tym możliwość manipulacji środowiskowej. Podobnie nielinearność gier, ich dynamiczny charakter oraz atrakcyjna forma są atutami, które pozwalają

eksperymentatorom na większe pole manewru przy testowaniu intuicji etycznych ludzi. Projekty typu *Moral Machine* są krokiem we właściwą stronę, sądzimy jednak, że ich upodobnienie do gier, przy zachowaniu wyjściowej idei badawczej, może pomóc w dokonywaniu ważnych odkryć w tej dziedzinie.

Bibliografia

Alexander, L., Moore, M. (2016). Deontological Ethics. W: E.N. Zalta (red.), *The Stanford encyclopedia of philosophy* (Winter 2016).

Blascovich, J., Loomis, J., Beall, A.C., Swinth, K.R., Hoyt, C.L., & Bailenson, J.N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2): 103–124.

Bonnefon, J.F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293): 1573–1576.

Bobrowski, M., Rodzińska-Szary, P., Socha, M. (2015). *Kondycja polskiej branży gier wideo. Raport 2015*. Dostępny online: http://kreatywna-europa.eu/wp-content/uploads/2016/01/Raport-na-temat-kondycji-polskiej-bran%C5%BCy-gier-wideo-1–1.pdf [data dostępu: 16.08.2017].

Cushman, F. (2013). Action, outcome, and value a dual-system framework for morality. *Personality and Social Psychology Review*, 17: 273–292.

DeSimone, J.A., Harms, P.D., & DeSimone, A.J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2): 171–181.

ESA - Entertainment Software Association (2016). *Essential facts about the computer and video game industry*. Dostępny online: http://essentialfacts.theesa.com/mobile/ [data dostępu: 16.08.2017].

Floridi, L. (2013). *The ethics of information*. Oxford University Press.

Frey, A., Hartig, J., Ketzel, A., Zinkernagel, A., & Moosbrugger, H. (2007). The use of virtual environments based on a modification of the computer game Quake III Arena® in psychological experimenting. *Computers in Human Behavior*, *23*(4): 2026–2039.

Gitelman, L. (red.). (2013). *Raw data is an oxymoron*. MIT Press.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., and Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44: 389–400.

Huebner, B., Lee, J.J., & Hauser, M.D. (2010). The moral-conventional distinction in mature moral competence. *Journal of Cognition and Culture*, *10*(1): 1–26.

Hursthouse, R., Pettigrove, G. (2016). Virtue Ethics. W: E.N. Zalta (red.), *The Stanford encyclopedia of philosophy* (Fall 2016).

Illari, P., & Russo, F. (2014). *Causality: Philosophical theory meets scientific practice*. OUP Oxford.

Järvelä, S., Ekman, I., Kivikangas, J.M., & Ravaja, N. (2012). Digital games as experiment stimulus. *Proceedings of DiGRA Nordic*, s. 6–8.

Johansson, R., & Nilsson, J. (2016). Disarming the trolley problem – why self-driving cars do not need to choose whom to kill. W: *Workshop CARS 2016 – Critical Automotive Applications: Robustness and Safety*, red. M. Roy (Goteborg). Dostępny online: https://hal.archives-ouvertes.fr/hal-01375606/file/CARS2016_paper_16.pdf [data dostępu: 16.08.2017].

Johansson-Stenman, O., & Martinsson, P. (2008). Are some lives more valuable? An ethical preferences approach. *Journal of health economics*, *27*(3): 739–752.

Kivikangas, J.M., Chanel, G., Cowley, B., Ekman, I., Salminen, M., Järvelä, S., & Ravaja, N. (2011). A review of the use of psychophysiological methods in game research. *Journal of gaming & virtual worlds*, *3*(3): 181–199.

Miller, B., & Record, I. (2013). Justified belief in a digital age: On the epistemic implications of secret Internet technologies. *Episteme*, *10*(2): 117–134.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2).

NHTSA (2013). Preliminary Statement of Policy Concerning Automated Vehicles. Dostępny online: https://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf [data dostępu: 16.08.2017].

Paxton, J.M., Ungar, L., & Greene, J.D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, 36: 163–177.

Sicart, M. (2009). The banality of simulated evil: designing ethical gameplay. *Ethics and information technology*, *11*(3): 191–202.

Sicart, M. (2013). Moral dilemmas in computer games. *Design Issues*, *29*(3): 28–37.

Sinnott-Armstrong, W. (2015). Consequentialism. W: E.N. Zalta (red.), *The Stanford encyclopedia of philosophy* (Winter 2015).

Sütfeld, L.R., Gast, R., König, P., & Pipa, G. (2017). Using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure. *Frontiers in behavioral neuroscience*, *11*.

Wallach, W., & Allen, C. (2008). *Moral machines: Teaching robots right from wrong*. Oxford University Press.

Washburn, D.A. (2003). The games psychologists play (and the data they provide). *Behavior Research Methods*, *35*(2): 185–193.

Ludografia

Bethesda Game Studios (2008). *Fallout 3* [PC]. USA: Bethesda Softworks.

BioWare (1998). *Baldur's Gate* [PC]. USA: Interplay Entertainment.

BioWare (2000). *Baldur's Gate II: Shadows of Amn* [PC]. USA: Interplay Entertainment.

CCP Games (2003). *Eve Online* [PC Computer, Online Game]. CCP Games: played 12 September 2011.

Lionhead Studios (2004). *Fable* [PC]. Wielka Brytania: Microsoft Game Studios.

Media Molecule (2008). *LittleBigPlanet* [PS3]. USA: Sony Computer Entertainment.

Mojang (2011). *Minecraft* [PC]. USA: Mojang.

Telltale Games (2012). *The Walking Dead* [PC]. USA: Telltale Games.

Valve Corporation, 1999 (mod) / 2000 (wersja pudełkowa), *Counter-Strike* [PC], USA: Vivendi.

Abstrakt

Celem artykułu jest zidentyfikowanie potencjalnych korzyści oraz zagrożeń związanych z badaniem dylematów moralnych dotyczących autonomicznych pojazdów za pomocą gier i zgamifikowanych narzędzi badawczych. Gry komputerowe dają naszym zdaniem możliwość skonstruowania bardziej ekologiczne trafnych eksperymentów. Dzięki zwiększonej trafności, wyniki eksperymentów mogą lepiej odzwierciedlać rzeczywiste mechanizmy podejmowania decyzji moralnych. Dane w nich uzyskane mogą posłużyć w implementacji „modułów moralnych" w bezzałogowych pojazdach. W artykule analizujemy wybrane eksperynty wykorzystujace elementy gier komputerowych. Analizy te zestawiamy z dostępnymi w literaturze podejściami do problemu algorytmizacji etyki oraz implementacji moralności w autonomicznych maszynach. Pokazujemy również, jak można wpisać ten problem w bardziej ogólny schemat „etyki informacji" zaproponowany przez Floridiego.

Słowa kluczowe: gamifikacja, intuicje etyczne, testowalność, algorytmizacja, bezzałogowe samochody, etyka informacji

Abstract

The aim of this paper is to identify potential dangers and benefits of investigating moral intuitions about the autonomous vehicles (AVs), using gamified research tools. We argue that computer games facilitate constructing more ecologically valid experiments. Due to the increased validity, the experimental outcomes can represent the real-life mechanisms of decision-making more faithfully. Data extracted in this manner could be used in implementation of 'moral modules' in AVs built in the real world. In our paper we analyze selected experimental setups that use game-like elements and assess them in the light of current literature regarding the algorithmization of ethics and implementability of ethical frameworks in machines. We also show how this problem can be interpreted in an overarching conceptual scheme of Floridi's 'information ethics'.

Keywords: gamification, moral intuitions, testability, algorythmisation, autonomous vehicles, information ethics

Autorzy

Bartosz Maćkiewicz (ur. 1992; b.mackiewicz@uw.edu.pl) – filozof, prawnik. Doktorant w Zakładzie Epistemologii UW, absolwent MISH i WPiA UW. Specjalizuje się w filozofii eksperymentalnej, filozofii nauki i zastosowaniu metod komputerowych w filozofii i prawie. Redaktor naukowy „Filozofii Nauki".

Wojciech Mamak (ur. 1992) – filozof i historyk nauki, absolwent MISH UW. Zajmuje się kognitywistyką, historią i filozofią nauki oraz epistemologią historyczną. Przygotowuje książkę dotyczącą rozwoju metod obliczeniowych i ich wpływu na zmiany ideałów racjonalności w Polsce w XIX i XX w.